

## Overview of the RNA G-quadruplex structures

Magdalena Malgowska, Karolina Czajczynska, Dorota Gudanis, Aleksander Tworak and Zofia Gdaniec✉

Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznań, Poland

**G-quadruplexes are non-canonical secondary structures which may be formed by guanine rich sequences, both *in vitro* and in living cells. The number of biological functions assigned to these structural motifs has grown rapidly since the discovery of their involvement in the telomere maintenance. Knowledge of the G-quadruplexes' three-dimensional structures plays an important role in understanding of their conformational diversity, physiological functions, and in the design of novel drugs targeting the G-quadruplexes. In the last decades, structural studies have been mainly focused on the DNA G-quadruplexes. Their RNA counterparts gained an increased interest along with a still-emerging recognition of the central role of RNA in multiple cellular processes. In this review we focus on structural properties of the RNA G-quadruplexes, based on high-resolution structures available in the RCSB PDB data base and on structural models. In addition, we point out the current challenges in this field of research.**

**Key words:** G-tetrad, U-tetrad, RNA G-quadruplex, NMR, crystallography, nucleic acids structure

**Received:** 30 May, 2016; **revised:** 28 June, 2016; **accepted:** 07 July, 2016; **available on-line:** 02 November, 2016

### INTRODUCTION

G-quadruplex is a structural motif formed by guanine-rich DNA and RNA molecules or nucleic acids analogues, such as LNA or PNA (Krishnan-Ghosh *et al.*, 2004; Randazzo *et al.*, 2004; Burge *et al.*, 2006; Ji *et al.*, 2011; Russo Krauss *et al.*, 2014). Most of the studied G-quadruplex structures are formed by a single type of the nucleic acid, however nucleic acids of a different type (e.g. DNA and RNA) may also interact to adopt a G-quadruplex fold (Xu *et al.*, 2009; Shrestha *et al.*, 2014).

The expansion of our knowledge on the G-quadruplex structures has become a priority in the last two decades in the medical, biological and chemical research for a number of reasons. Structures of this type were identified in multiple regions of various viral, prokaryotic and eukaryotic genomes (Schaffitzel *et al.*, 2001; Cahoon & Seifert, 2009; Shen *et al.*, 2011), as well as multiple cellular RNAs (Kumari *et al.*, 2007; Millevoi *et al.*, 2012; Agarwala *et al.*, 2015). Recently, an *in vivo* existence of the DNA and RNA G-quadruplexes has been unequivocally proven (Biffi *et al.*, 2013; Biffi *et al.*, 2014). The observed abundance of those structural motifs and their retention over a long evolutionary period suggest their involvement in multiple biological processes.

One of the first functional studies in which the formation of an RNA G-quadruplex was coupled with a particular biological function concerned the 127-nt fragment

derived from the 5' end of the HIV *gag* gene (Sundquist & Heaphy 1993). Multiple lines of evidence available to date suggest that the dimerization of the two copies of HIV-1 genomic RNA is mediated by the formation of an intermolecular G-quadruplex structure (Awang & Sen, 1993; Marquet *et al.*, 1994; Shen *et al.*, 2011). This dimerization process is known to influence multiple stages of the viral life cycle. Additional studies have revealed that the G-quadruplex structures are present in both, the HIV-1 genomic RNA and its DNA copy, and may therefore represent attractive targets for antiviral interventions (Metifiot *et al.*, 2014). In 2007 Azzalin and co-workers have demonstrated that the mammalian telomeric DNA is transcribed into long non-coding RNAs (known as TERRA RNAs) composed of tandem repeats of the UUAGGG motif (Azzalin *et al.*, 2007). At that time, the corresponding short DNA motif was already known to form a G-quadruplex structure (Wang & Patel, 1993; Parkinson *et al.*, 2002; Gavathiotis & Searle, 2003; Zhang *et al.*, 2005; Ambrus *et al.*, 2006). Thus, the discovery of TERRA RNAs gained a broad interest in the occurrence of the RNA G-quadruplex structures in living cells and their possible biological roles. Indeed, shortly after, RNAs composed of the tandem UUAGGG repeats were shown to adopt a stable G-quadruplex structure (Xu *et al.*, 2008a; Xu *et al.*, 2008b; Kimura *et al.*, 2009).

Subsequently, numerous studies have shown that the RNA G-quadruplexes are involved in various biological processes, including translation regulation, mRNA processing, transcription termination, mRNA localization, as well as alternative splicing. Most commonly, presence of a G-quadruplex structure within the transcript results in either up- or down-regulation of the gene from which it originates. Although expression of a number of genes was shown to be altered by the presence of this structural motif within the mRNA molecule, at present little is known about the mechanisms underlying these processes. The G-quadruplex structure itself may impede the movement of the pre-initiation complex or the ribosome along the mRNA molecule. Alternatively, a G-quadruplex may recruit protein factors which influence the transcript in different ways. For a detailed review of known biological roles of the RNA G-quadruplexes see (Agarwala *et al.*, 2015). Owing to their broad distribution and function, the RNA G-quadruplexes quickly became

✉ e-mail: [zgdan@ibch.poznan.pl](mailto:zgdan@ibch.poznan.pl)

**Abbreviations:** <sup>8</sup>G, 8-bromoguanosine; <sup>5</sup>U, 5-bromouridine; CD, Circular Dichroism; d(<sup>8</sup>G), 8-bromo-2'-deoxyguanosine; d(<sup>5</sup>U), 5-bromo-2'-deoxyuridine; DLS, Dynamic Light Scattering; HIV, Human Immunodeficiency Virus; LNA, Locked Nucleic Acids; NMR, Nuclear Magnetic Resonance; PNA, Peptide Nucleic Acids; RCSB PDB, Research Collaboratory for Structural Bioinformatics, Protein Data Bank; SAXS, Small-angle X-ray scattering; TERRA, Telomeric Repeat-containing RNAs.

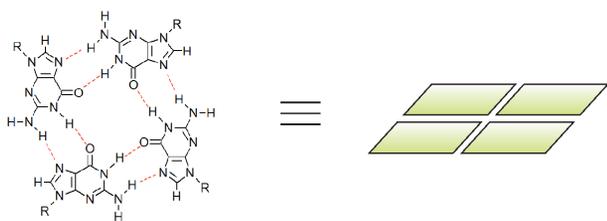


Figure 1. Schematic representation of a G-tetrad.

an interesting target for novel drug development strategies, including new antiviral, anticancer and neurological disease therapies (Collie & Parkinson, 2011; Ji *et al.*, 2011; Xu & Komiyama, 2012; Murat *et al.*, 2014). For example, various small molecules have been successfully employed to regulate the translation process *via* trapping or disrupting of the G-quadruplex motifs found in multiple cellular mRNAs (Bugaut *et al.*, 2010; Gomez *et al.*, 2010; Halder *et al.*, 2011; Faudale *et al.*, 2012; Morris *et al.*, 2012). The challenge in this field is to find the ligands which will exhibit high specificity towards the RNA G-quadruplex targets with minimal or none off-target effects. Recently, an RNA G-quadruplex database, G4RNA, was created by Garant and co-workers to facilitate systematic research and comparative analyses (Garant *et al.*, 2015).

The fundamental structural unit of the G-quadruplex is a guanosine-tetrad (G-tetrad, Fig. 1). In a G-tetrad, each of the four guanine bases is a donor of two hydrogen bonds at the Watson-Crick edge and an acceptor of two hydrogen bonds at the Hoogsteen edge. Four carbonyl oxygen atoms, directed to the center of the G-tetrad, form a negatively charged central core of a G-quadruplex. In order to compensate for the negative charge and to stabilize the G-quadruplex structure, the presence of monovalent cations, in particular  $K^+$  and to a lesser extent  $NH_4^+$  and  $Na^+$ , is required. When the ionic radius of the cation is relatively small (e.g. in the case of  $Na^+$ ), the ions can be located in the middle of a single G-tetrad and coordinate four oxygen atoms. Alternatively, the ion may be embedded between two G-tetrads and coordinate eight oxygen atoms, which is the case for some larger cations, like  $K^+$  or  $NH_4^+$ .

Depending on the nucleic acid strand orientation, a G-quadruplex can be described as parallel or anti-parallel (Fig. 2). A parallel G-quadruplex is formed when all four strands are oriented in the same direction. As a consequence, all residues found in each individual G-tetrad have the same glycosidic bond conformation (*anti-anti-*

*anti-anti* or *syn-syn-syn-syn*). Antiparallel G-quadruplexes are formed when at least one DNA/RNA strand runs in an orientation opposite to the other strands. They can be further divided into two different types, depending on the number of oppositely oriented strands. The first type of antiparallel G-quadruplexes is formed when two strands run in the opposite orientation to the other two, and the G-tetrads comprise two guanines in the *syn* and two in the *anti* conformations (*syn-syn-anti-anti* and *syn-anti-syn-anti*). The second type of antiparallel G-quadruplex structures is known as a hybrid-type (3+1) and occurs when three strands are parallel to each other, while the fourth strand runs in the opposite direction (Fig. 2). In the (3+1) G-quadruplex, three guanines adopt the same, and the fourth residue exhibits an opposite conformation (*syn-anti-anti-anti* and *anti-syn-syn-syn*).

In general, G-quadruplexes may be comprised of one strand with at least four G-rich fragments, folded to adopt an intramolecular structure. Alternatively, they may be formed by two or four separate strands. It should be emphasized that, to date, all known naturally occurring RNA G-quadruplexes are parallel stranded, with only one exception: an unusual G-quadruplex motif found in a 97-nt RNA fluorogenic aptamer named Spinach (Huang *et al.*, 2014; Warner *et al.*, 2014). The strong preference of the RNA G-quadruplexes to adopt the parallel fold is a consequence of the strong propensity of the riboguanosine for the *anti* conformation. The partially locked conformation around the glycosidic bond is due to the steric hindrance caused by the presence of a C2' hydroxyl group which forces the C3'-*endo* ribose conformation.

Within the G-quadruplex structure, the G-tetrads may be connected by loops of variable size and sequence. Based on their arrangement within the structure, three major loop types may be distinguished: (i) a double chain reversal loop (also known as a propeller) which connects the adjacent parallel strands, (ii) a diagonal loop connecting two opposite antiparallel strands, and (iii) an edgewise loop (also known as a lateral loop) which connects two adjacent antiparallel strands (Fig. 3). All G-quadruplex structures have four grooves, whose dimensions (either narrow, medium or wide) are defined by the glycosidic conformation of the adjacent bases in the tetrad. A more detailed insight into the fundamentals of the G-quadruplex structure can be found in an excellent review by Burge and co-workers (Burge *et al.*, 2006).

In contrast to the functional studies, the structural studies of the RNA G-quadruplexes were performed less intensively, focusing mainly on the DNA molecules. In consequence, only 18 RNA G-quadruplex

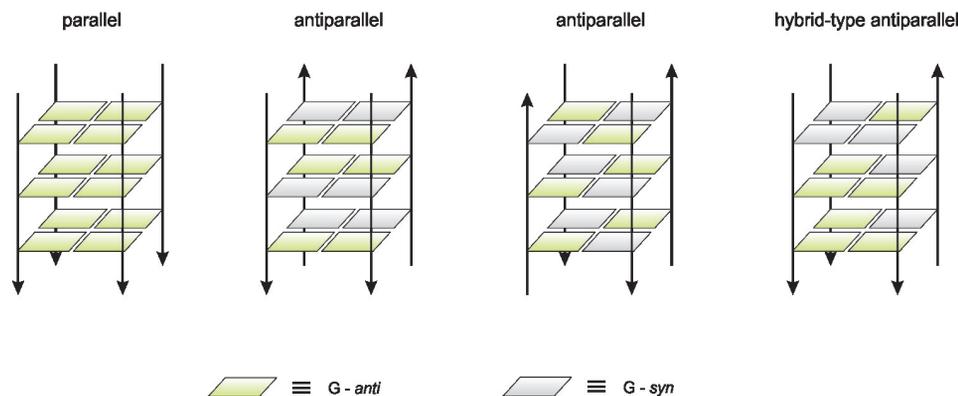
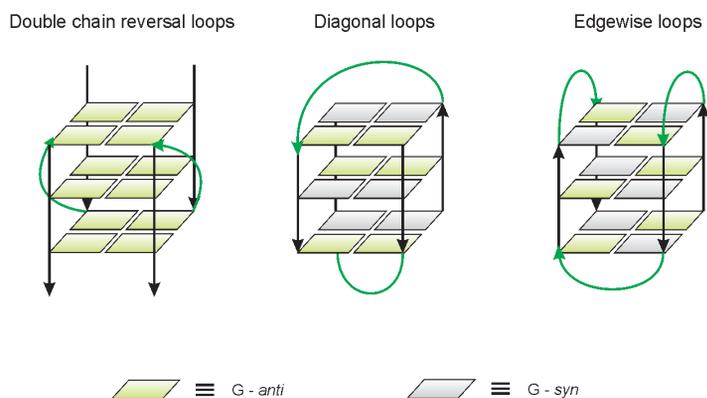


Figure 2. Strand polarities in parallel, antiparallel and hybrid-type antiparallel G-quadruplexes together with a corresponding *syn-anti* orientation of guanosine residues. Strand polarities are indicated as the arrows.



**Figure 3. Examples of the loop arrangements in inter- and intramolecular G-quadruplexes.**

The loops are marked in green.

structures are currently available in the RCSB PDB database, including 12 X-ray structures and 6 NMR-derived structures (Table 1). Although knowledge on existing topologies of the DNA and RNA G-quadruplexes has been growing constantly, a recent analysis by Sket and Plavec (Sket & Plavec, 2015) suggests that it is impossible to predict the G-quadruplex folding topology based solely on the oligonucleotide sequence. This is because the final fold of a given G-quadruplex depends on multiple factors, such as the number of guanosine residues in the sequence, the length and the composition of the nucleic acid sequences connecting the G-rich fragments, the *syn/anti* conformation around the glycosidic bonds, the presence of chemically modified nucleotides, the DNA/RNA concentration and the nature of the ions stabilizing the structure. Moreover, analysis of the already determined high resolution structures of the DNA and RNA G-quadruplexes revealed their great diversity and conformational polymorphism. For example, it was shown that the same d(AGGGTTAGGGTTAGGGTTAGGG) molecule may adopt two different G-quadruplex folds depending on the environmental conditions. Its NMR solution structure obtained in the presence of Na<sup>+</sup> ions is an antiparallel G-quadruplex with one diagonal and two lateral loops (Wang & Patel, 1993), while in the crystal, a parallel G-quadruplex structure with three propeller loops was determined in the presence of K<sup>+</sup> ions (Parkinson *et al.*, 2002).

#### METHODS FOR STUDYING A 3D G-QUADRUPLEX STRUCTURE

Many experimental methods are used for structural studies of the G-quadruplexes. Among them, NMR spectroscopy and X-ray crystallography are the only methods which enable determination of the three-dimensional structures at atomic resolution.

The unique feature of the NMR spectroscopy is a possibility of carrying out experiments under close to physiological conditions. There are, however, certain limitations of this method. The major one is a requirement for the presence of a single conformation in solution, which gives sharp and well resolved resonances. Another limitation is that the samples need to be very pure and relatively highly concentrated (preferably over 1 mM). Yet, the increased concentration may lead to aggregation

of the molecules, causing further problems in the spectrum interpretation. Details concerning the peaks assignment strategy, fold determination and calculation of the G-quadruplex structures can be found in two great review articles (Webba da Silva, 2007; Adrian *et al.*, 2012).

It should be kept in mind that even a simple 1D <sup>1</sup>H NMR spectrum may provide useful information on the G-quadruplex conformation in the solution. This is due to the fact that imino proton resonances of the G-tetrads appear in a characteristic spectral region: between 10–12 ppm. This spectral region, typical for Hoogsteen-type interactions, is separated from the imino region, where imino protons from the Watson-Crick base pairs resonate (13–14 ppm). Typically, the number of imino proton resonances observed in the 10–12 ppm region corresponds to the number of guanosine residues involved in the G-tetrad formation, unless the structure is symmetric.

Another characteristic feature of the imino proton resonances originating from the hydrogen atoms involved in the G-tetrad formation, especially those originating from the G-tetrads hidden inside a G-quadruplex structure, is that they can be detected even if the spectra are recorded at high temperatures, sometimes over 90°C. Some of those resonances do not exchange for deuterium when the solvent is changed from H<sub>2</sub>O to D<sub>2</sub>O and can be detected after many hours, days, or even months.

X-ray crystallography is the second method which allows to determine the 3D G-quadruplex structures with the highest resolution. X-ray crystallography has proven to be uniquely informative, since it can provide not only the three-dimensional arrangement of atoms but also information on cation binding sites and hydration patterns. The biggest challenge of this method is to obtain well diffracting single crystals. For this reason, various modifications within the G-quadruplex forming DNA/RNA molecules are often introduced, and multiple crystallization conditions are screened. High purity and homogeneity of the sample are crucial for the crystallization to be successful. Similarly to the NMR study, high concentrations of nucleic acids are required for crystallization experiments (e.g., 1 to 2 mM of single-stranded DNA/RNA). A detailed report on the crystallization methodology for G-quadruplexes and their complexes with ligands has been provided by Campbell and co-workers (Campbell & Parkinson, 2007; Campbell *et al.*, 2012).

SAXS and DLS are low resolution methods used to study the size and overall shape of the G-quadruplexes in solution. It was shown, using both techniques, that the 5' guanosine monophosphate in a solution containing potassium ions tends to self-assemble into aggregates composed of stacking G-tetrad layers (Spindler *et al.*, 2004; Mariani *et al.*, 2009; Mariani *et al.*, 2010). In fact, NMR structural data can be supplemented by experimental information from SAXS, to validate the proposed structural models (Gudanis *et al.*, 2016).

#### RNA G-QUADRUPLEX STRUCTURES

The first structure of an RNA G-quadruplex at the atomic resolution was published in 1992 (Cheong & Moore, 1992). Based on the analysis of the NMR spectra recorded in the presence of 50 mM potassium ions, it was shown that four strands of r(UGGGGU) form

a helical, parallel-stranded G-quadruplex (PDB ID: 1RAU). In this structure, four G-tetrads (all guanine residues in the *anti* conformation) were flanked by two uridine-tetrads (U-tetrads), one on each end of the structure (Fig. 4a). U-tetrads were formed by four uracil residues connected through a cyclic hydrogen-bond pattern, which comes from the interaction between N<sub>3</sub>H imino hydrogen atoms of one uridine and O4 carbonyl oxygen atoms of the adjacent bases (Fig. 5a). In the r(UGGGGU) structure, the U-tetrad formed by the four 3' uridines appeared to be very stable, its imino proton resonance (U6 N3H) were observed even at 65°C. The presence of a U-tetrad at the 5' end has been proposed based on the observation of the additional uridine imino resonance at a temperature below 25°C that was shifted downfield in relation to a free imino proton resonance.

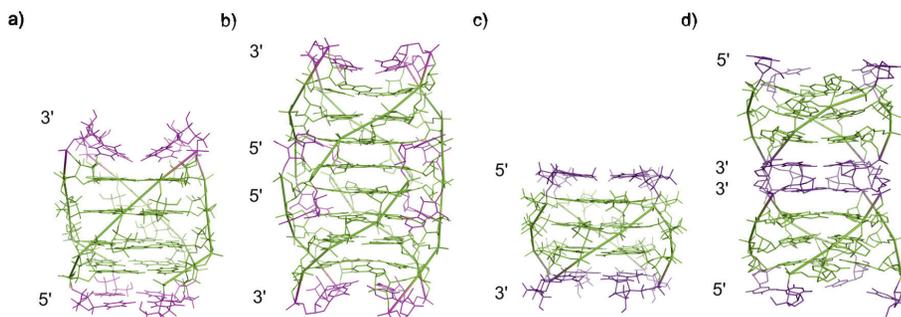
Additional insight into the structural diversity of the RNA G-quadruplexes emerged from the crystal structure of the same r(UGGGGU) molecules, solved at subatomic resolution (0.61 Å) for crystals grown in the presence of Sr<sup>2+</sup> ions (PDB ID: 1J8G) (Deng *et al.*, 2001). Similarly to the previously described solution structure, r(UGGGGU) in the crystalline state forms a parallel four-stranded G-quadruplex with all guanines exhibiting *anti* conformation, and Sr<sup>2+</sup> sandwiched between every G-tetrad plane. In this structure, two G-quadruplexes form 5'-5' intertwined dimers. The dimer structure includes six G-tetrads and two U-tetrads (one at each end of the dimer, both composed of the 3' uridines). In turn, all 5' uridines stack out, rotate around the backbone of G2, and form G:U:G:U:G:U:G:U octad motifs (Fig. 5b) with the neighboring G-quadruplexes (Fig. 4b).

Recently, Fyfe and co-workers (Fyfe *et al.*, 2015) further examined the r(UGGGGU) structure stabilized by Sr<sup>2+</sup> ions, using X-ray crystallography and solution SAXS analysis. The aim of this study was to elucidate factors that contribute to the high stability of the RNA G-quadruplexes, much higher than their relative DNA counterparts. It was shown that networks of water-mediated contacts within the helical grooves of RNA G-quadruplexes greatly enhance their overall stability. Comparison of atomic coordinates of the four r(UGGGGU) G-quadruplex crystal structures coming from this study (PDB IDs: 1J8G, 4RKV, 4RJ1, 4RNE) revealed a considerable conformational variability of the 3' U-tetrad. In addition, the SAXS-derived solution structure of r(UGGGGU) was in a good agreement with the dimeric assembly observed in the crystal structure, however, differing from the conformation observed by NMR spectroscopy (Cheong & Moore, 1992). This difference was likely related to the type of coordinated cation, Sr<sup>2+</sup> *versus* K<sup>+</sup>.

The LNAs are nucleotide analogues that are locked in a C3'-*endo* sugar conformation by a methylene bridge between 2'-O and 4'-C atoms in the ribose, which leads to an *anti* conformation of the glycosidic torsion angle and reduces the overall conformational flexibility. This locked conformation resembles the conformation of RNA residues, thus it was not surprising that the l(TGGGT), in a solution containing potassium ions, folds into a G-quadruplex structure (NMR study, PDB ID: 1S9L; Randazzo *et al.*, 2004) very similar to the one determined by Cheong and Moore for r(UGGGGU) (Cheong & Moore, 1992), as shown in Fig. 4a and c. The main difference between these two solution structures concerns the presence or absence of the terminal tetrads. In contrast to the NMR spectrum of r(UGGGGU), no thymine imino resonances were observed in the l(TGGGT) spectrum. On the other hand, the presence of the T-tetrad (Fig. 5c) was observed in the crystal structure of l(TGGGT) G-quadruplex (Fig. 4d). This 3' T-tetrad is stabilized by cyclic N3H – O4 hydrogen bonds and by a perfect stacking of the bases on the six-membered rings of the preceding G-tetrad (PDB ID: 4L0A; Russo Krauss *et al.*, 2014). Interestingly, two such parallel G-quadruplexes interact *via* stacking of the two terminal 3' T-tetrads to form an elongated dimer. By this observation, a novel, opposite polarity (tail-to-tail) arrangement was reported, as only head-to-head and head-to-tail interfaces had been shown previously in the crystal structures of G-quadruplexes (Russo Krauss *et al.*, 2014).

A comparative analysis of the described helical, parallel-stranded tetramolecular RNA G-quadruplexes allows to draw several conclusions on their general structural properties. Typically, in a solution containing potassium ions, G-quadruplexes of this type occur in a monomeric form, while crystallization probably enhances formation of dimers and other highly ordered arrangements. The terminal uridine residues can form U-tetrads which are always more stable when formed by 3' uridines, than 5' Us. In the RNA G-quadruplex crystal structure, the U-tetrad plane seems quite distorted, and the four uridines are tilted. In contrast, the T-tetrad, if present in the LNA G-quadruplex structure, is very regular and perfectly stacks on the six-membered rings of the preceding G-tetrad.

Another group of RNA G-quadruplexes, extensively studied because of their great functional potential, originates from the human telomeric repeats. As mentioned before, the possibility of a G-quadruplex formation by the TERRA molecules was first demonstrated by Xu and co-workers (Xu *et al.*, 2008b). Based on the analysis of the NMR spectra recorded for r(UAGGGUUAGGGU), i.e. two repeats of the telomeric motif, in the pres-



**Figure 4. Comparison of the structures of:** a) r(UGGGGU), 1RAU; b) r(UGGGGU), 1J8G; c) l(TGGGT), 1S9L and d) l(TGGGT), 4L0A. The residues are marked as follow: guanines are green, uridines – magenta, and LNA thymidines – dark purple.

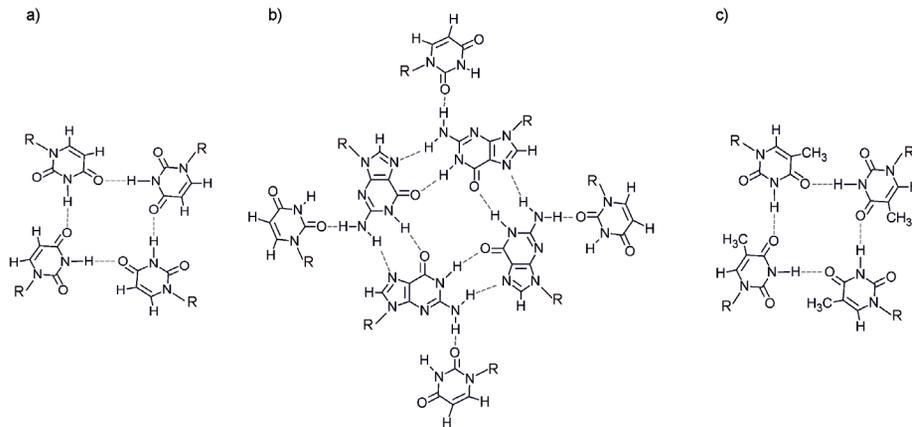


Figure 5. Schematic representation of: a) U-tetrad, b) G:U:G:U:G:U octad and c) T-tetrad.

ence of sodium ions, the authors proposed a bimolecular, parallel stranded G-quadruplex structural model. The model was composed of three G-tetrads with two double-chain-reversal UUA loops. Almost at the same time, a high resolution NMR structure in potassium solution was published for the same TERRA molecule, r(UAGGGUUAGGGU), by Martadinata & Phan (PDB ID: 2KBP; Martadinata & Phan, 2009). This structure showed good agreement with the previously described model (Fig. 6a). The observed propeller-type parallel-stranded RNA G-quadruplex involved all guanosine residues in the *anti* conformation. Two double-chain-reversal UUA loops were located at the medium-size grooves, linking two adjacent parallel strands and bridging three G-tetrad layers. The apparent similarity between the two RNA structures stays in high contrast to the results reported for human telomeric DNA sequences for which various G-quadruplex topologies were observed in the presence of potassium and sodium ions (Martadinata & Phan, 2009). This structural conservation of the RNA G-quadruplex structures was attributed to the preference of the riboguanosine to adopt the *anti* conformation.

Martadinata and Phan complemented their study with structural analysis of two truncated variants of the previously described 12-nt TERRA RNA. They analyzed 10-nt r(GGGUUAGGGU) and 9-nt r(GGGUUAGGG) using gel electrophoresis, CD and NMR spectroscopy, and observed dimerization of bimolecular G-quadruplexes in both cases. Additionally, 9-nt RNA seemed to adopt some higher-order structures. These results suggested that the presence of the 5' UA and 3' U flanking residues

prevented aggregation of the r(UAGGGUUAGGGU) G-quadruplexes (Martadinata & Phan, 2009).

High-resolution structure of the 10-nt TERRA was determined a few years later by the same researchers (PDB ID: 2M18) (Martadinata & Phan, 2013). The solution structure in the presence of potassium ions revealed many structural details of the stacking interface and loop conformations. Two G-quadruplex units were stacked on their 5' ends (Fig. 6b). The stacking induced considerable loop rearrangements in comparison with the previously reported 12-nt TERRA structure. Adenine bases in each UUA loop were positioned coplanar with the 5' G-tetrad. The N1 nitrogen atoms of two adenosines were hydrogen-bonded to 2'-OH groups of the adjacent guanosine residues forming a specific A:(G-G-G-G)·A hexad (Fig. 7a). The presence of this hexad led to increased stacking contacts between the two G-quadruplex units and contributed to the enhanced stabilization of the stacking conformation. The same stacking interaction and loop conformation was reported in 2011 by Collie and co-workers for a unique high-resolution structure of TERRA in a complex with a small-molecule ligand (PDB ID: 3MIJ), which will be described in the next section (Collie *et al.*, 2011).

One year earlier, the same group managed to solve the crystal structure of a modified variant of the previously described 12-nt TERRA sequence, r(<sup>B</sup>UAGGGUUAGGGU) (PDB ID: 3IBK; Collie *et al.*, 2010). It was in good agreement with the previously described NMR structure of the unmodified RNA (Martadinata & Phan, 2009). Both G-quadruplex structures are

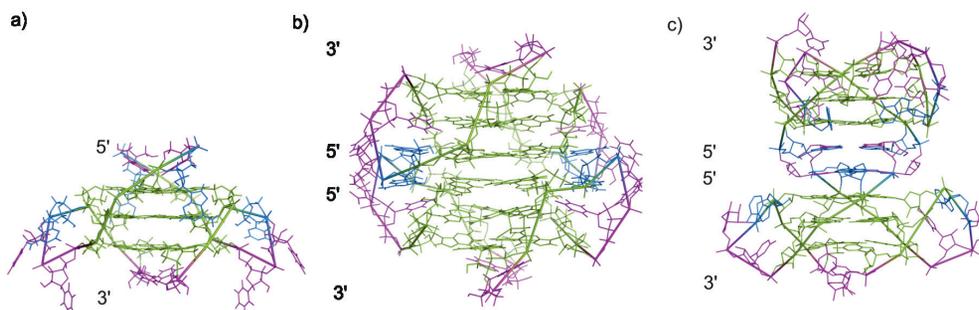


Figure 6. The comparison of two TERRA repeats structures: a) r(UAGGGUUAGGGU), 2KBP; b) r(GGGUUAGGGU), 2M18; c) r(<sup>B</sup>UAGGGUUAGGGU), 3IBK.

The residues are marked as follow: guanosines are green, uridines – magenta, and adenosines – blue.

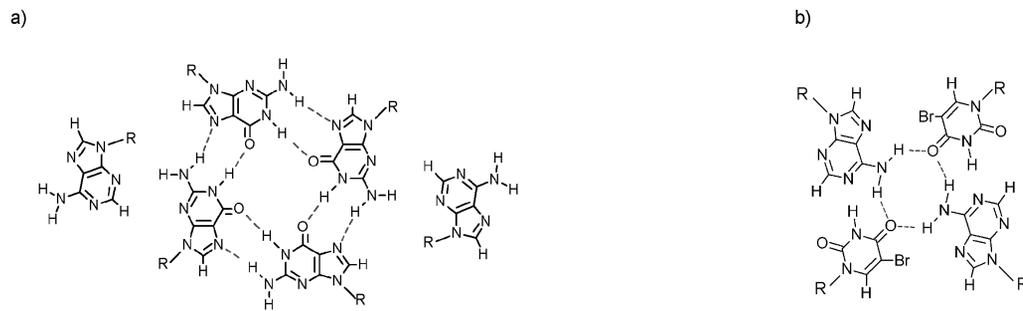


Figure 7. Schematic representation of: a) A-(G-G-G-G)-A hexad and b) BrU:A:BrU:A tetrad occurring in TERRA G-quadruplexes

bimolecular, parallel-stranded, and possess two double-chain-reversal UUA loops. However, in the crystal structure, the two G-quadruplex units were interlocked with each other at their 5' ends. This interaction included a formation of two stacked  ${}^{\text{Br}}\text{U}:\text{A}:\text{BrU}:\text{A}$  tetrads between the two G-quadruplex units, involving the first two residues of each RNA strand (Fig. 6c and 7b).

It was shown that RNA composed of only one telomeric repeat, r(UAGGGU), is also capable of forming a G-quadruplex structure (Xu *et al.*, 2010). The detailed analysis of NMR spectra recorded in the presence of sodium and potassium ions revealed that r(UAGGGU) forms a tetramolecular, parallel-stranded G-quadruplex, which is capped and highly stabilized by a 3' U-tetrad (Fig. 8). In addition, presence of the adenosine residue seemed to hinder the 5' U-tetrad formation.

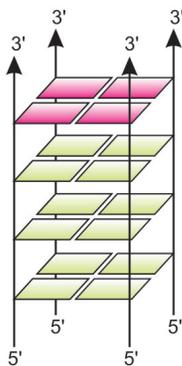


Figure 8. The model of tetramolecular parallel r(UAGGGU) G-quadruplex (Xu *et al.*, 2010). G-tetrads are marked in green and the 3' U-tetrad is shown in magenta.

Martadinata and co-workers investigated physiologically relevant long TERRA sequences (of up to 96-nt in length) and addressed a question about the arrangement of the G-quadruplex blocks along these molecules (Martadinata *et al.*, 2011). A molecular dynamics simulations study on an eight-repeat sequence which is long enough to form two consecutive G-quadruplex blocks connected by the UUA linker, was performed for the four possible stacking modes: head-to-head (5'-5'), tail-to-tail (3'-3'), tail-to-head (3'-5') and head-to-tail (5'-3'), as shown in Fig. 9. Only the head-to-tail arrangement was not stable throughout the simulations, while in the head-to-head interaction, the contiguous stacking of the loop bases and loop-loop interactions between two subunits were observed. Based on these results, the authors have proposed a structural model for long human TERRA molecules, in which the G-quadruplexes are arranged in a “beads-on-a-string” fashion, i.e. they can move relatively independently of each other and are constrained only by the connecting linkers (Fig. 10).

Another distinct group of RNA sequences known to adopt a G-quadruplex fold contains the GGAGG stretches. A common feature of these structures is the presence of A-(G-G-G-G)-A hexad and formation of dimers through the hexad-hexad stacking (Liu *et al.*, 2002; Lipay & Mihailescu, 2009; Mashima *et al.*, 2009; Medic *et al.*, 2014; Malgowska *et al.*, 2014). The A-(G-G-G-G)-A hexad differs from the one formed by the 10-nt TERRA sequence described above. In this hexad, two adenosine residues interact with a canonical G-tetrad. Their N6H amino hydrogens and N7 nitrogen atoms are hydrogen-bonded to guanosine N3 nitrogen atom and N2H amino hydrogen, respectively (Fig. 11).

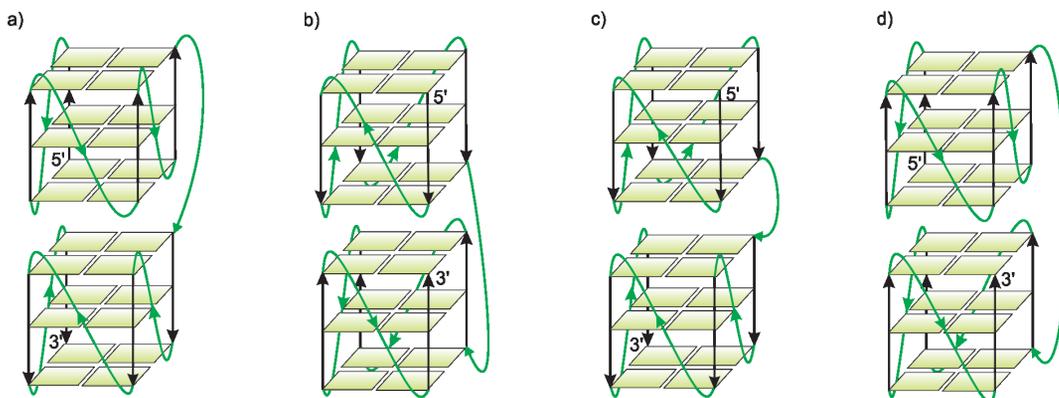


Figure 9. Schemes of the possible stacking modes between two consecutive TERRA G-quadruplex units. a) 5'-5'; b) 3'-3'; c) 3'-5'; d) 5'-3'. The UUA loops are colored in dark green.

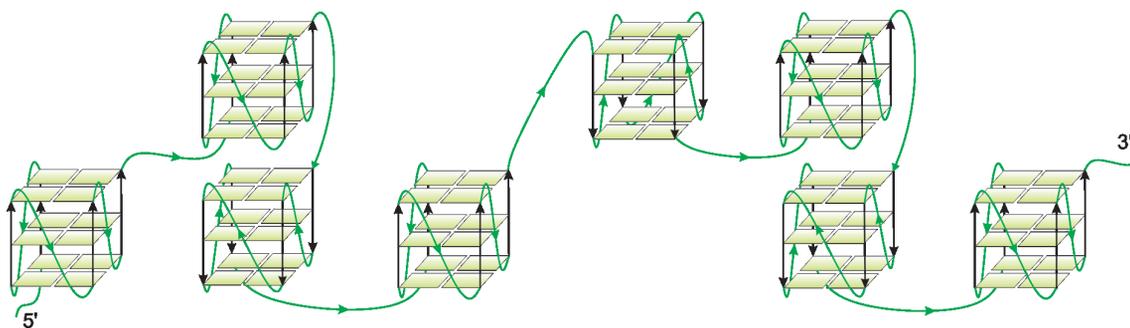


Figure 10. A schematic model of a long human TERRA arranged as “beads-on-a-string”. The UUA loops and linkers are colored in dark green.

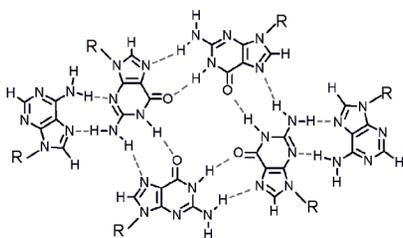


Figure 11. Schematic representation of an A-(G-G-G-G)-A hexad.

Three sequence variants of RNA molecules containing a single GGAGG motif: UGGAGGU, GGAGGA and AGGAGGA, are known to preferentially fold into bimolecular G-quadruplexes with the characteristic A-(G-G-G-G)-A hexad motif. NMR studies conducted in solutions containing potassium ions have shown that, indeed, all the structures associate by stacking through the interface of two hexads (Lipay & Mihailescu, 2009; Medic *et al.*, 2014; Malgowska *et al.*, 2014). This structural model is schematically shown in Fig. 12.

In addition, two RNA G-quadruplex structures formed by molecules comprising two GGAGG segments are available in the RSCB PDB database (PDB entries: 1MY9, 2RQJ). In both cases the structure was determined by NMR spectroscopy in the presence of potassium ions. In the first structure (Liu *et al.*, 2002), RNA with two GGAGG stretches joined by a UUUU segment, r(GGAGGUUUUGGAGG), folds into an intramolecular parallel G-quadruplex. Analysis revealed presence of a G-tetrad, UUUU double-chain reversal loop, and A-(G-G-G-G)-A hexad. Furthermore, two G-quadruplex units dimerize through the head-to-head stacking interactions, which occur between the interfaces of the two hexads (Fig. 13a). A similar type of RNA dimer is formed by r(GGAGGAGGAGGA) (PDB ID: 2RQJ; Mashima *et al.*, 2009), suggesting that even a single nucleotide linker between the two GGAGG motifs allows the G-quadruplex structure formation (Fig. 13b).

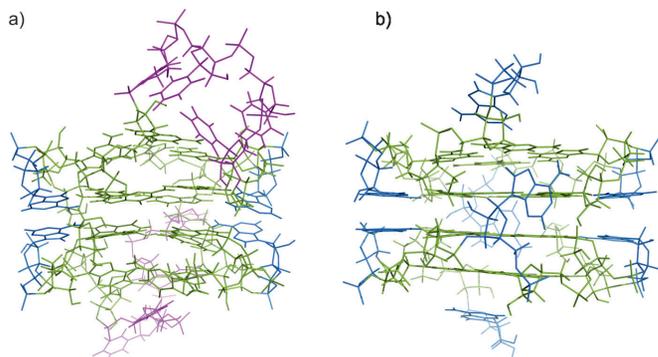


Figure 13. Comparison of G-quadruplex structures adopted by sequences comprising two r(GGAGG) motifs: a) r(GGAGGUUUUGGAGG), 1MY9; and b) r(GGAGGAGGAGGA), 2RQJ.

Guanosine residues are shown in green, uridines in magenta, and adenines in blue.

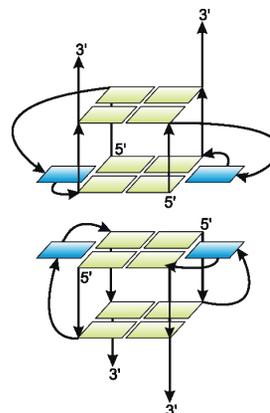
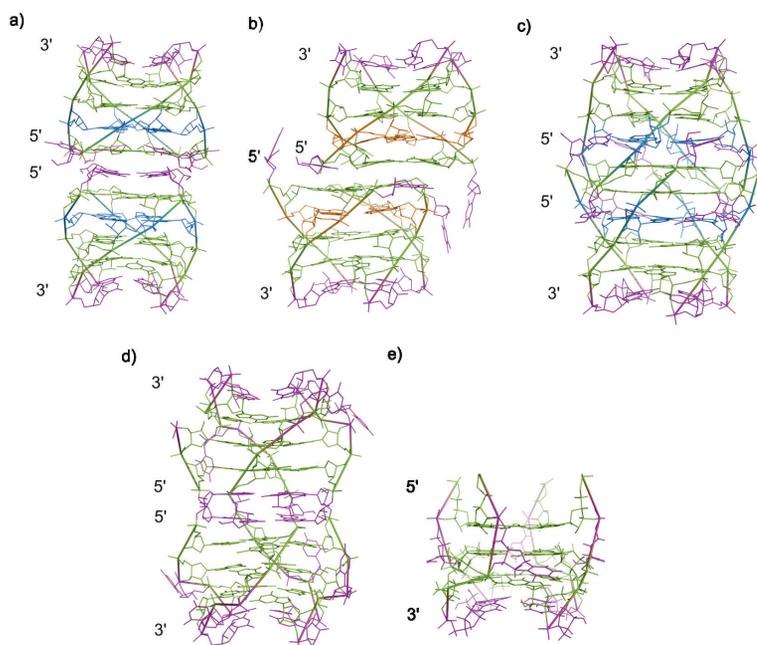
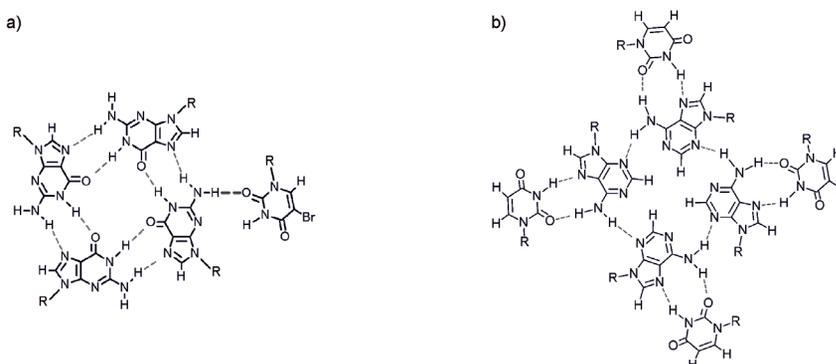


Figure 12. Dimerization of bimolecular G-quadruplexes formed by short sequences comprising one r(GGAGG) motif. G-tetrads are marked in green and the adenine residues participating in the A-(G-G-G-G)-A hexad formation are shown in blue.

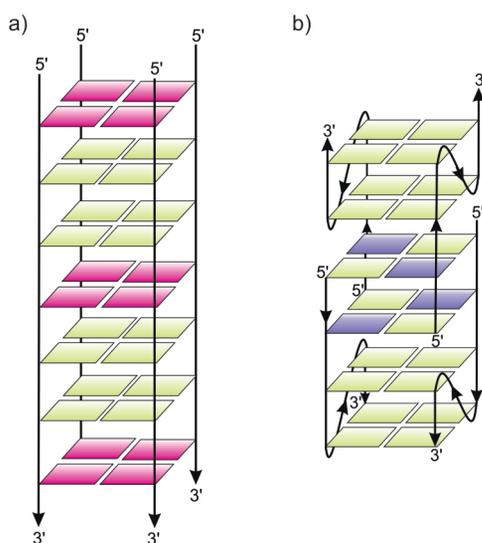
Another interesting class of the RNA G-quadruplexes is represented by a group of the crystal structures determined in the Sundaralingam laboratory (Pan *et al.*, 2003; Pan *et al.*, 2003a; Pan *et al.*, 2003b; Pan *et al.*, 2006a; Pan *et al.*, 2006b). Analysis of these structures has revealed a number of novel and unexpected features concerning the tetramolecular G-quadruplexes. In general, the overall conformation of these RNAs with comparable sequences, rUd<sup>(Br)G</sup>r(AGGU), rUd<sup>(Br)G</sup>r(UGGU), rUd<sup>(Br)G</sup>r(GUGU), d<sup>(Br)U</sup>r(GAGGU) and d<sup>(Br)U</sup>r(GIGGU) is similar. All fold into tetramolecular, parallel stranded G-quadruplexes which, with the exception of rUd<sup>(Br)G</sup>r(UGGU), dimerize in a head-to-head alignment. All of the guanosine residues form G-tetrads and all the 3' terminal uridines form buckled U-tetrads stabilized by one cyclic N3H – O4 hydrogen bond. The 5' terminus of these structures shows a variety of interaction patterns: stacking d<sup>(Br)U</sup>r(GAGGU), d<sup>(Br)U</sup>r(GIGGU), groove binding rUd<sup>(Br)G</sup>r(AGGU), or intercalation r(U)d<sup>(Br)G</sup>r(GUGU). In d<sup>(Br)U</sup>r(GAGGU) structure (PDB ID: 1J6S), the stacking interface involves a 5' d<sup>(Br)U</sup>-tetrad formed by one subunit of the dimer and a G:U:G:U:G:U:G:U octad coming from the second subunit, which contributes to the in-



**Figure 14.** Comparison of X-ray structures of: a)  $d^{(Br)U}r(GAGGU)$ , 1J6S; b)  $d^{(Br)U}r(GIGGU)$ , 2GRB; c)  $rUd^{(Br)G}r(AGGU)$ , 1MDG; d)  $r(U)d^{(Br)G}r(GUGU)$ , 2AWE; e)  $rUd^{(Br)G}r(UGGU)$ , 1P79. The guanines residues are shown in green, uridines in magenta, adenosines in blue and inosines in orange.



**Figure 15.** Schematic representation of: a)  $U:(G:G:G:G)$  pentad, and b)  $A:U:A:U:A:U$  octad.



creased stacking surface between the two units (Fig. 14a). The octad is formed by four 5'  $d^{(Br)U}$  that swing out of the G-quadruplex core and interact with the preceding G-tetrad by hydrogen bonding. Stacking of the symmetric  $G:U:G:U:G:U:G:U$  octad results in the co-axial packing of two G-quadruplex units (Pan *et al.*, 2003). In contrast, a non-co-axial packing mode within a dimer of two G-quadruplexes was observed in the crystal structure of  $d^{(Br)U}r(GIGGU)$  (PDB ID: 2GRB; Fig. 14b). Stacking of two asymmetric  $U:(G:G:G:G)$  pentads (Fig. 15a), found at the 5' ends of two G-quadruplex units, causes the displacement of two G-quadruplex axes by about 7.4 Å (Pan *et al.*, 2006b).

The groove binding pattern of interaction of the two G-quadruplex subunits was observed in the crystal structure of  $rUd^{(Br)G}r(AGGU)$  (PDB ID: 1MDG). In this dimeric structure, the adenine residues form

**Figure 16.** The models of G-quadruplexes composed of two UGG and CGG repeats: a)  $r(UGGUGGU)$  and b)  $r(GCGGCGGC)$ . G-tetrads are marked in green and the U-tetrads are shown in magenta and cytidine residues are colored in purple.

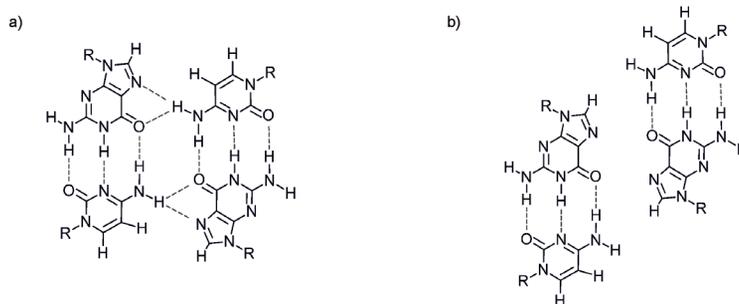


Figure 17. Schematic representation of G:C:G:C tetrads from: a) r(GCGGCGGC), and b) r(GC(BrG)GCGGC) G-quadruplexes.

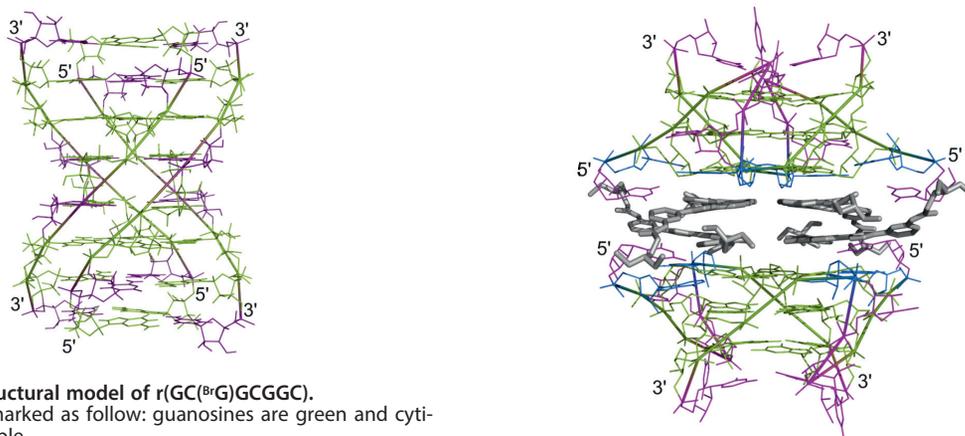


Figure 18. The structural model of r(GC(BrG)GCGGC).

The residues are marked as follow: guanosines are green and cytosines are dark purple.

two A-tetrads and further interact in the reverse Hoogsteen pairing scheme with the 5' uridines to form two A:U:A:U:A:U:A:U octads (Fig. 14c and 15b). In addition to the groove binding interaction, this dimer is stabilized by intercalation of the G-tetrads immediately adjacent to the 5' uridines (Pan *et al.*, 2003a). Another example of the intercalation-type interaction resulting in the dimerization process is represented by the structure of r(U)d(BrG)r(GUGU) (PDB ID: 2AWE). In this case, swapping of the 5' U-tetrads was observed (Fig. 14d), as well as a bulged-out conformation of the internal uridine bases (Pan *et al.*, 2006a). The rUd(BrG)r(UGGU) G-quadruplex (PDB ID: 1P79) was also shown to possess the bulged-out U motif (Fig. 14e; Pan *et al.*, 2003b), which indicates that substitution of purine (A, G or I) with uridine may lead to a bulged-out conformation of the pyrimidine within a G-quadruplex structure.

The potential of RNAs composed of two and four AGG, CGG or UGG trinucleotide repeats to form G-quadruplex structures is intensively studied in our laboratory (Malgowska *et al.*, 2014; Gudanis *et al.*, 2016). Expansion of those trinucleotide motifs was linked with many neurological disorders. In general, we have shown that all of these molecules can fold into the G-quadruplex

Figure 19. The structure of the r(UAGGGUUAGGGU) – acridine complex, 3MIJ.

Guanosine residues are indicated in green, uridines in magenta, adenosines in blue and two acridine ligand molecules in grey.

plex structures. However, in case of the CGG-repeat RNAs, the G-quadruplex formation competes with their strong tendency to form duplex and hairpin structures. We also demonstrated that in a solution containing potassium ions all oligoribonucleotides comprising two trinucleotide repeats fold preferentially into bimolecular G-quadruplexes with distinct structural motifs which are characteristic of the type of repeats.

Our recent study has shown that RNA which contains two AGG repeats, r(AGGAGGA), forms a dimer of bimolecular G-quadruplexes. The interaction between the two subunits involves stacking of two A:(G·G·G·G)·A hexads. The observed structure appears to be similar to the previously described RNA G-quadruplex containing the GGAGG stretches (Fig. 12). All experimental data indicated that r(UGGUGGU) adopts a symmetrical tetramolecular parallel G-quadruplex structure with four G-tetrads and three U-tetrads (Fig. 5a, 16a), the 3' U-tetrad being the most stable. Although there is no known

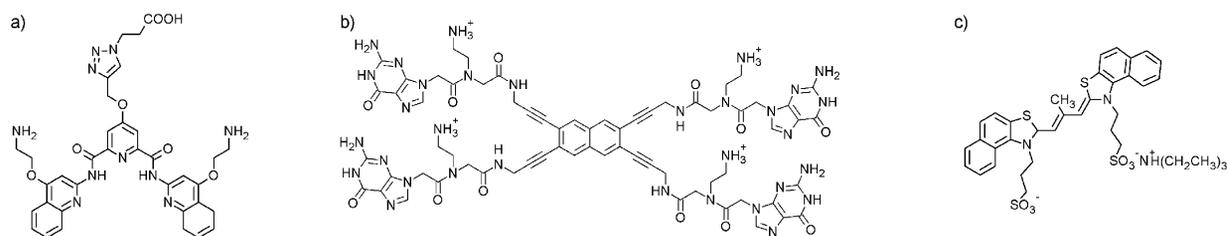


Figure 20. Chemical structure of the RNA G-quadruplex-binding ligands: a) carboxyPDS, b) NaphthoTASQ, c) CyT.

Table 1. Summary of the published RNA G-quadruplex structures.

Sequence	PDB ID	Type	Topology	NMR/X-ray	Reference
r(UGGGGU)	1RAU	Tetramolecular	Parallel	NMR	Cheong and Moore 1992
r(UGGGGU)	1J8G	Dimer of tetramolecular G-quadruplexes	Parallel	X-ray	J. Deng, Xiong, and Sundaralingam 2001
r(UGGGGU)	4RKV 4RJ1 4RNE	Tetramolecular	Parallel	X-ray	Fyfe et al. 2015
l(TGGGT)	1S9L	Tetramolecular	Parallel	NMR	Randazzo et al. 2004
l(TGGGT)	4L0A	Tetramolecular	Parallel	X-ray	Russo Krauss et al. 2014
r(UAGGGUUAGGGU)	2KBP	Bimolecular	Parallel	NMR	Martadinata and Phan 2009
r(GGGUUAGGGU)	2M18	Dimer of bimolecular G-quadruplexes	Parallel	NMR	Martadinata and Phan 2013
r(UAGGGUUAGGGU)	3MIJ	Bimolecular	Parallel	X-ray	Collie et al. 2011
r( <sup>Br</sup> UAGGGUUAGGGU)	3IBK	Dimer of bimolecular G-quadruplexes	Parallel	X-ray	Collie et al. 2010
r(GGAGGUUUUGGAGG)	1MY9	Bimolecular	Parallel	NMR	Liu et al. 2002
r(GGAGGAGGAGGA)	2RQJ	Bimolecular	Parallel	NMR	Mashima et al. 2009
d( <sup>Br</sup> U)r(GAGGU)	1J6S	Tetramolecular	Parallel	X-ray	Pan et al. 2003
d( <sup>Br</sup> U)r(GIGGU)	2GRB	Tetramolecular	Parallel	X-ray	Pan et al. 2006b
r(U)d( <sup>Br</sup> G)r(AGGU)	1MDG	Dimer of tetramolecular G-quadruplexes	Parallel	X-ray	Pan et al. 2003a
r(U)d( <sup>Br</sup> G)r(GUGU)	2AWE	Dimer of tetramolecular G-quadruplexes	Parallel	X-ray	Pan et al. 2006a
r(U)d( <sup>Br</sup> G)r(UGGU)	1P79	Tetramolecular	Parallel	X-ray	Pan et al. 2003b

RNA G-quadruplex structure with an internal U-tetrad, the presence of imino resonance assigned to the internal uridine residue, which was observed even at a high temperature, indicated the formation of a stable U-tetrad accommodated in the center of the G-quadruplex. The 5' U-tetrad was less stable and its imino proton was observed only below 25°C (Malgowska *et al.*, 2014).

Very interesting results were also obtained for molecules composed of two repeats of the CGG motif. In case of r(GCGGCGGC), a tetramolecular, parallel, highly symmetrical, interlocked G-quadruplex dimer was found (Fig. 16b). It included four G-tetrads and two mixed G:C:G:C tetrads (Fig. 17a). However, when one of the guanosine residues was replaced by the 8-bromoguanosine, r(GC(<sup>Br</sup>G)GCGGC), a totally novel, tetramolecular, antiparallel architecture of the RNA G-quadruplexes was obtained (Fig. 18). In our study, we have provided evidence that in solution, an antiparallel RNA G-quadruplex is formed by the self-association of two duplexes *via* the major groove. In this G-quadruplex, two <sup>Br</sup>G3:G6:<sup>Br</sup>G3:G6 tetrads are sandwiched between mixed G:C:G:C tetrads (Fig. 17b). The antiparallel fold is facilitated by the presence of an 8-bromoguanosine residue, which initially promotes formation of a stable <sup>Br</sup>G:G mismatch and subsequently induces the formation of <sup>Br</sup>G3:G6:<sup>Br</sup>G3:G6 tetrads (Gudanis *et al.*, 2016).

#### CHEMICAL TARGETING OF RNA G-QUADRUPLEXES

Our expanding knowledge on the functional role of the RNA G-quadruplexes in living cells raises the need for highly selective molecular tools that bind to these structures with high affinity. This is a goal of high importance especially in the field of medical research (Kerwin, 2000; Le *et al.*, 2012; Biffi *et al.*, 2013; Biffi *et al.*,

2014; Xiong *et al.*, 2015). So far, a wide range of small-molecule ligands has been identified to exhibit selectivity for G-quadruplexes over single- or double stranded nucleic acids (Zhang *et al.*, 2014). Many of these small molecules comprise heterocyclic ring systems, including anthraquinones, pyridostatin, acridines, perylenes and porphyrins (Vummidi *et al.*, 2013), which are capable of interacting with the terminal G-tetrads of RNA G-quadruplex structures. However, most of these compounds have been discovered by targeted screening of small-molecule libraries, rather than by rational design. This is due to a very limited availability of structural data for the G-quadruplex-ligand complexes. Moreover, the available structures are dominated by crystallographic studies of complexes in which DNA molecules were used. Only one high-resolution RNA G-quadruplex-ligand structure has been reported to date. The ligand belongs to the family of 3,6-disubstituted acridines and was shown to bind the TERRA molecules (Collie *et al.*, 2011). In the crystal structure of this complex (PDB ID: 3MIJ), two ligand molecules stack side-by-side on the 5' G-tetrad of a bimolecular, parallel-stranded G-quadruplex. Two such G-quadruplex subunits interact in a head-to-head manner, by this way sandwiching two layers of ligand molecules (Fig. 19). Ligand binding induces a substantial reorganization of the G-quadruplex loop conformation. The adenine residues from the UUA reversal loops and the incomplete 5' UA loops align in the 5' G-tetrad plane and form a G:A:G:A:G:A:G:A octad. In consequence, the surface available for the  $\pi$ - $\pi$  ligand stacking becomes markedly extended. The formation of this platform is stabilized by multiple hydrogen bonds with the 2'-OH groups of the UUA loops. The ligand itself is hydrogen-bonded to the RNA molecules. A weak hydrogen bond between the amide carbonyl group of the

ligand side chain and the N3 nitrogen of the 5' U has been determined.

In comparison to an analogous DNA complex, some significant differences occur. The DNA G-quadruplex complex with the acridine ligand (PDB ID: 3QCR) has a 1:1 stoichiometry with a single ligand molecule bound on the 5' G-tetrad plane. In contrast, the TERRA-acridine ligand complex (PDB ID: 3MIJ) has a 2:1 stoichiometry of acridine molecule recognition. Furthermore, there are no loop-acridine contacts in the DNA complexes and no G:A:G:A:G:A:G:A octad. Most importantly, the conformational change within the TERRA loops after ligand binding suggests that rational design of ligands specific towards certain G-quadruplexes cannot be solely based on their native structures (Collie *et al.*, 2011).

The structural similarity between scaffolds of the parallel DNA and RNA G-quadruplexes makes it challenging to design small molecules that provide recognition selectivity only for the RNA. However, recently it was shown that a small molecule, carboxypyridostatin (carboxyPDS) (Fig. 20a), selectively stabilizes the RNA over the DNA G-quadruplexes (Di Antonio *et al.*, 2012). This unique property may allow for selective identification and visualization of the RNA G-quadruplexes *in vivo*. Indeed, it was shown that carboxyPDS can selectively target and trap the RNA G-quadruplexes *versus* DNA G-quadruplex structures within a cellular context (Biffi *et al.*, 2014). Other ligands, NaphthoTASQ and CyT (Fig. 20b, c), were also proposed for detection and visualization of the RNA G-quadruplexes *in vivo* (Laguerre *et al.*, 2015; Xu *et al.*, 2015). Although both showed a very high selectivity for G-quadruplexes compared to other structural motifs, they were not able to discriminate between the DNA and RNA G-quadruplex variants.

The RNA G-quadruplex structures appear to be involved in a much wider range of biological processes than their DNA counterparts (Eddy & Maizels, 2007; Huppert *et al.*, 2008; Millevoi *et al.*, 2012). Thus, it is possible that research efforts will soon focus mainly on rational design of small molecules targeting the RNA G-quadruplexes (Balasubramanian & Neidle, 2009; Satyanarayana *et al.*, 2010; Morris *et al.*, 2012).

## CONCLUSIONS

The number of known DNA and RNA G-quadruplex structures is still very limited when compared to the large number of potential G-quadruplex-forming sequences predicted by bioinformatic analyses (Huppert & Balasubramanian, 2005; Todd *et al.*, 2005; Huppert *et al.*, 2008; Huppert, 2008). This observation strongly suggests that a large number of diverse G-quadruplex topologies is still unknown, waiting to be explored. In fact, a new RNA G-quadruplex motif of unusual topology has been identified very recently in an *in vitro* selected RNA aptamer (known as Spinach) that binds a GFP-like ligand and activates its green fluorescence (Huang *et al.*, 2014; Warner *et al.*, 2014). In this structure, a three-tetrad G-quadruplex core is composed of two G-tetrads stacked above a mixed-sequence tetrad. The tetrads are connected by five loops of various length which are composed from one up to 34-nt. The most unexpected feature of this G-quadruplex motif is its antiparallel topology that is unprecedented for RNA G-quadruplexes. We believe that multiple areas of research will greatly benefit from a growing number of high-resolution structures of the RNA G-quadruplexes. Together with better understanding of the G-quadruplex topologies they will contribute

to the identification of new G-quadruplex ligands as possible drug candidates for future clinical applications.

## Acknowledgements

This work was supported by the Ministry of Science and Higher Education of the Republic of Poland by the KNOW program and National Science Center (NCN) under Grant No. UMO-2014/13/B/ST5/04144.

## REFERENCES

- Adrian M, Heddi B, Phan AT (2012) NMR Spectroscopy of G-Quadruplexes. *Methods* **57**: 11–24. doi:10.1016/j.ymeth.2012.05.003
- Agarwala P, Pandey S, Maiti S (2015) The Tale of RNA G-Quadruplex. *Org Biomol Chem* **13**: 5570–5585. doi:10.1039/C4OB02681K
- Ambrus A, Chen D, Dai J, Bialis T, Jones R, Yang D (2006) Human telomeric sequence forms a hybrid-type intramolecular g-quadruplex structure with mixed parallel/antiparallel strands in potassium solution. *Nucleic Acids Res* **34**: 2723–2735. doi:10.1093/nar/gkl348
- Awang G, Sen D (1993) Mode of Dimerization of HIV-1 Genomic RNA. *Biochemistry* **32**: 11453–11457. doi:10.1021/bi00093a024
- Azzalin CM, Reichenbach P, Khoraiuli L, Giulotto E, Lingner J (2007) Telomeric repeat containing RNA and RNA surveillance factors at mammalian chromosome ends. *Science* **318**: 798–801. doi:10.1126/science.1147182
- Balasubramanian S, Neidle S (2009) G-Quadruplex nucleic acids as therapeutic targets. *Curr Opin Chem Biol* **13**: 345–353. doi:10.1016/j.cbpa.2009.04.637
- Biffi G, Tannahill D, McCafferty J, Balasubramanian S (2013) Quantitative visualization of DNA G-quadruplex structures in human cells. *Nature Chem* **5**: 182–186. doi:10.1038/nchem.1548
- Biffi G, Di Antonio M, Tannahill D, Balasubramanian S (2014) Visualization and selective chemical targeting of RNA G-quadruplex structures in the cytoplasm of human cells. *Nature Chem* **6**: 75–80. doi:10.1038/nchem.1805
- Bugaut A, Rodriguez R, Kumari S, Hsu STD, Balasubramanian S (2010) Small molecule-mediated inhibition of translation by targeting a native RNA G-quadruplex. *Organic & Biomol Chem* **8**: 2771–2776. doi:10.1039/c002418j
- Burge S, Parkinson GN, Hazel P, Todd AK, Neidle S (2006) Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res* **34**: 5402–5415. doi:10.1093/nar/gkl655
- Cahoon LA, Seifert HS (2009) An alternative DNA structure is necessary for pilin antigenic variation in neisseria gonorrhoeae. *Science* **325**: 764–767. doi:10.1126/science.1175653
- Campbell NH, Parkinson GN (2007) Crystallographic studies of quadruplex nucleic acids. *Methods* **43**: 252–263. doi:10.1016/j.ymeth.2007.08.005
- Campbell NH, Collie GW, Neidle S (2012) Crystallography of DNA and RNA G-quadruplex nucleic acids and their ligand complexes. *Curr Protocols in Nucleic Acid Chem*. Chapter 17: Unit17.6. doi:10.1002/0471142700.nc1706s50
- Cheong C, Moore PB (1992) Solution structure of an unusually stable RNA Tetraplex containing G- and U-tetrad structures. *Biochemistry* **31**: 8406–8414. doi:10.1021/bi00151a003
- Collie GW, Parkinson GN (2011) The application of DNA and RNA G-quadruplexes to therapeutic medicines. *Chem Soc Rev* **40**: 5867–5892. doi:10.1039/c1cs15067g
- Collie GW, Haider SM, Neidle S, Parkinson GN (2010) A crystallographic and modelling study of a human telomeric RNA (TERRA) quadruplex. *Nucleic Acids Res* **38**: 5569–5580. doi: 10.1093/nar/gkq259
- Collie GW, Sparapani S, Parkinson GN, Neidle S (2011) Structural basis of telomeric RNA quadruplex-acridine ligand recognition. *J Am Chem Soc* **133**: 2721–2728. doi:10.1021/ja109767y
- Deng J, Xiong Y, Sundaralingam M (2001) X-Ray analysis of an RNA tetraplex (UGGGU)<sub>4</sub> with divalent Sr<sup>2+</sup> ions at subatomic resolution (0.61 Å). *Proc Natl Acad Sci USA* **98**: 13665–13670. doi:10.1073/pnas.241374798
- Deng Z, Norseen J, Wiedmer A, Riethman H, Lieberman PM (2009) TERRA RNA binding to TRF2 facilitates heterochromatin formation and orc recruitment at telomeres. *Molecular Cell* **35**: 403–413. doi:10.1016/j.molcel.2009.06.025
- Di Antonio M, Biffi G, Mariani A, Raiber EA, Rodriguez R, Balasubramanian S (2012) Selective RNA *versus* DNA G-quadruplex targeting by *in situ* click chemistry. *Angewandte Chemie International Edition* **51**: 11073–11078. doi:10.1002/anie.201206281
- Eddy J, Maizels N (2007) Conserved elements with potential to form polymorphic G-quadruplex structures in the first intron of human genes. *Nucleic Acids Res* **36**: 1321–1333. doi:10.1093/nar/gkm1138
- Faudale M, Cogo S, Xodo LE (2012) Photoactivated cationic alkyl-substituted porphyrin binding to g4-RNA in the 5'-UTR of

- KRAS oncogene represses translation. *Chem Commun* **48**: 874–876. doi:10.1039/C1CC15850C
- Fyfe AC, Dunten PW, Martick MM, Scott WG (2015) Structural variations and solvent structure of r(UGGGGU) quadruplexes stabilized by Sr<sup>2+</sup> Ions. *J Mol Biol* **427**: 2205–2219. doi:10.1016/j.jmb.2015.03.022
- Garant JM, Luce MJ, Scott MS, Perreault JP (2015) G4RNA: An RNA G-quadruplex database. *Database* **2015**: 1–5. doi:10.1093/database/bav059
- Gavathiotis E, Searle MS (2003) Structure of the parallel-stranded DNA quadruplex d(TTAGGGT)<sub>4</sub> containing the human telomeric repeat: evidence for A-Tetrad formation from NMR and molecular dynamics simulations. *Organic & Biomol Chem* **1**: 1650–1656. doi:10.1039/b300845m
- Gomez D, Guedin A, Mergny J-L, Salles B, Riou J-F, Teulade-Fichou M-P, Calsou P (2010) A G-quadruplex structure within the 5'-UTR of TRF2 mRNA represses translation in human cells. *Nucleic Acids Res* **38**: 7187–7198. doi:10.1093/nar/gkq563
- Gudanis D, Popena L, Szpotkowski K, Kierzek R, Gdaniec Z (2016) Structural characterization of a dimer of RNA duplexes composed of 8-bromoguanosine modified CGG trinucleotide repeats: a novel architecture of RNA quadruplexes. *Nucleic Acids Res* **44**: 2409–2416. doi:10.1093/nar/gkv1534
- Halder K, Largy E, Benzler M, Teulade-Fichou MP, Hartig JS (2011) Efficient suppression of gene expression by targeting 5'-UTR-based RNA quadruplexes with bisquinolinium compounds. *Chem Bio Chem* **12**: 1663–1668. doi:10.1002/cbic.201100228
- Huang H, Suslov NB, Li NS, Shelke SA, Evans ME, Koldobskaya Y, Rice PA, Piccirilli JA (2014) A G-quadruplex-containing RNA activates fluorescence in a GFP-like fluorophore. *Nature Chem Biol* **10**: 686–691. doi:10.1038/nchembio.1561
- Huppert JL, Balasubramanian S (2005) Prevalence of quadruplexes in the human genome. *Nucleic Acids Res* **33**: 2908–2916. doi:10.1093/nar/gki609
- Huppert JL, Bugaut A, Kumari S, Balasubramanian S (2008) G-quadruplexes: the beginning and end of UTRs. *Nucleic Acids Res* **36**: 6260–6268. doi:10.1093/nar/gkn511
- Huppert JL (2008) Hunting G-quadruplexes. *Biochimie* **90**: 1140–1148. doi:10.1016/j.biochi.2008.01.014
- Ji X, Sun H, Zhou H, Xiang J, Tang Y, Zhao C (2011) Research progress of RNA quadruplex. *Nucleic Acid Therapeutics* **21**: 185–200. doi:10.1089/nat.2010.0272
- Kerwin S (2000) G-quadruplex DNA as a target for drug design. *Curr Pharmaceutical Design* **6**: 441–471. doi:10.2174/1381612003400849
- Kimura T, Xu Y, Komiya M (2009) Human Telomeric RNA r(UAGGGU) Sequence forms parallel tetraplex structure with U-quartet. *Nucleic Acids Symposium Series* **53**: 239–240. doi:10.1093/nass/nrp120
- Krishnan-Ghosh Y, Stephens E, Balasubramanian S (2004) A PNA<sub>4</sub> quadruplex. *J Am Chem Soc* **126**: 5944–5945. doi:10.1021/ja031508f
- Kumari S, Bugaut A, Huppert JL, Balasubramanian S (2007) An RNA G-quadruplex in the 5' UTR of the NRAS proto-oncogene modulates translation. *Nature Chem Biol* **3**: 218–221. doi:10.1038/nchembio864
- Laguette A, Hukezalie K, Winckler P, Katranji F, Chanteloup G, Pirrotta M, Perrier-Cornet JM, Wong JMY, Monchard D (2015) Visualization of RNA-quadruplexes in live cells. *J Am Chem Soc* **137**: 8521–8525. doi:10.1021/jacs.5b03413
- Le TVT, Han S, Chae J, Park HJ (2012) G-quadruplex binding ligands: from naturally occurring to rationally designed molecules. *Curr Pharmaceutical Design* **18**: 1948–1972. doi:10.2174/138161212799958431
- Lipay JM, Mihailescu MR (2009) NMR spectroscopy and kinetic studies of the quadruplex forming RNA r(UGGAGGU). *Mol Bio Systems* **5**: 1347–1355. doi:10.1039/b911555b
- Liu H, Matsugami A, Katahira M, Uesugi S (2002) A dimeric RNA quadruplex architecture comprised of two G:G(:A):G:G(:A) hexads, G:G:G:G tetrads and UUUU loops. *J Mol Biol* **322**: 955–970. doi:10.1016/S0022-2836(02)00876-8
- Malgowska M, Gudanis D, Kierzek R, Wyszko E, Gabelica V, Gdaniec Z (2014) Distinctive structural motifs of RNA G-quadruplexes composed of AGG, CGG and UGG trinucleotide repeats. *Nucleic Acids Res* **42**: 10196–10207. doi:10.1093/nar/gku710
- Mariani P, Spinozzi F, Federiconi F, Amenitsch H, Spindler L, Drevensk-Olenik I (2009) Small angle X-ray scattering analysis of deoxyguanosine 5'-monophosphate self-assembling in solution: nucleation and growth of G-quadruplexes. *J Phys Chem B* **113**: 7934–44. doi:10.1021/jp809734p
- Mariani P, Spinozzi F, Federiconi F, Ortore MG, Amenitsch H, Spindler L, Drevensk-Olenik I (2010) Guanosine quadruplexes in solution: a small-angle X-ray scattering analysis of temperature effects on self-assembling of deoxyguanosine monophosphate. *J Nucleic Acids* **2010**: 1–10. doi:10.4061/2010/472478
- Marquet R, Paillart JC, Skripkin E, Ehresmann C, Ehresmann B (1994) Dimerization of human immunodeficiency virus type 1 RNA involves sequences located upstream of the splice donor site. *Nucleic Acids Res* **22**: 145–151. doi:10.1093/nar/22.2.145
- Martadinata H, Phan AT (2009) Structure of propeller-type parallel-stranded RNA G-quadruplexes, formed by human telomeric RNA sequences in K<sup>+</sup> solution. *J Am Chem Soc* **131**: 2570–2578. doi:10.1021/ja806592z
- Martadinata H, Heddi B, Lim KW, Phan AT (2011) Structure of long human telomeric RNA (TERRA): G-quadruplexes formed by four and eight UUAGGG repeats are stable building blocks. *Biochemistry* **50**: 6455–6461. doi:10.1021/bi200569f
- Martadinata H, Phan AT (2013) Structure of human telomeric RNA (TERRA): stacking of two G-quadruplex blocks in K<sup>+</sup> solution. *Biochemistry* **52**: 2176–21783. doi:10.1021/bi301606u
- Mashima T, Matsugami A, Nishikawa F, Nishikawa S, Katahira M (2009) Unique quadruplex structure and interaction of an RNA aptamer against bovine prion protein. *Nucleic Acids Res* **37**: 6249–6258. doi:10.1093/nar/gkp647
- Medic Š, Podbevšek P, Plavec J (2014) Solution structure of a prion protein aptamer analogue. *Croatia Chemica Acta* **87**: 321–325. doi:10.5562/cca2430
- Metifiot M, Amrane S, Litvak S, Andreola ML (2014) G-quadruplexes in viruses: function and potential therapeutic applications. *Nucleic Acids Res* **42**: 12352–12366. doi:10.1093/nar/gku999
- Millevoi, Stefania, Hervé Moine, and Stéphan Vagner. (2012) G-quadruplexes in RNA biology. *Wiley Interdisciplinary Reviews. RNA* **3**: 495–507. doi:10.1002/wrna.1113
- Morris MJ, Wingate KL, Silwal J, Leeper TC, Basu S (2012) The porphyrin TmPyP4 unfolds the extremely stable G-quadruplex in M13-MMP mRNA and alleviates its repressive effect to enhance translation in eukaryotic cells. *Nucleic Acids Res* **40**: 4137–4145. doi:10.1093/nar/gkr1308
- Murat P, Zhong J, Lekiéffre L, Cowieson NP, Clancy JL, Preiss T, Balasubramanian S, Khanna R, Tellam J (2014) G-quadruplexes regulate Epstein-Barr virus-encoded nuclear antigen 1 mRNA translation. *Nature Chem Biol* **10**: 358–364. doi:10.1038/nchembio.1479
- Nandakumar J, Podell ER, Cech TR (2010) How telomeric protein POT1 avoids RNA to achieve specificity for single-stranded DNA. *Proc Natl Acad Sci USA* **107**: 651–656. doi:10.1073/pnas.0911099107
- Pan B, Xiong Y, Shi K, Deng J, Sundaralingam M (2003) Crystal structure of an RNA purine-rich tetraplex containing adenine tetrads. *Structure* **11**: 815–823. doi:10.1016/S0969-2126(03)00107-2
- Pan B, Xiong Y, Shi K, Sundaralingam M (2003a) An eight-stranded helical fragment in RNA crystal structure. *Structure* **11**: 825–831. doi:10.1016/S0969-2126(03)00108-4
- Pan B, Xiong Y, Shi K, Sundaralingam M (2003b) Crystal structure of a bulged RNA tetraplex at 1.1 Å resolution. *Structure* **11**: 1423–1430. doi:10.1016/j.str.2003.09.017
- Pan B, Shi K, Sundaralingam M (2006a) Base-tetrad swapping results in dimerization of RNA quadruplexes: implications for formation of the 1-motif RNA octaplex. *Proc Natl Acad Sci USA* **103**: 3130–3134. doi:10.1073/pnas.0507730103
- Pan B, Shi K, Sundaralingam M (2006b) Crystal structure of an RNA quadruplex containing inosine tetrad: implications for the roles of NH<sub>2</sub> group in purine tetrads. *J Mol Biol* **363**: 451–459. doi:10.1016/j.jmb.2006.08.022
- Parkinson GN, Lee MP, Neidle S (2002) Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature* **417**: 876–880. doi:10.1038/nature755
- Randazzo A, Esposito V, Ohlenschläger O, Ramachandran R, Mayola L (2004) NMR solution structure of a parallel LNA quadruplex. *Nucleic Acids Res* **32**: 3083–3092. doi:10.1093/nar/gkh629
- Redon S, Reichenbach P, Lingner J (2010) The non-coding RNA TERRA is a natural ligand and direct inhibitor of human telomerase. *Nucleic Acids Res* **38**: 5797–5806. doi:10.1093/nar/gkq296
- Russo Krauss I, Parkinson GR, Merlino A, Mattia CA, Randazzo A, Novellino E, Mazzarella L, Sica F (2014) A regular thymine tetrad and a peculiar supramolecular assembly in the first crystal structure of an All-LNA G-quadruplex. *Acta Crystallographica. Section D, Biological Crystallography* **70**: 362–370. doi:10.1107/S1399004713028095
- Satyanarayana M, Kim YA, Rzuczek SG, Pilch DS, Liu AA, Liu LF, Rice JE, LaVoie EJ (2010) Macrocyclic hexaaxozoles: influence of aminoalkyl substituents on RNA and DNA G-quadruplex stabilization and cytotoxicity. *Bioorganic & Med Chem Lett* **20**: 3150–3154. doi:10.1016/j.bmcl.2010.03.086
- Schaffitzel C, Berger I, Postberg J, Hanes J, Lipps HJ, Plückthun A (2001) *In vitro* generated antibodies specific for telomeric guanine-quadruplex DNA react with stylonychia lemnae macronuclei. *Proc Natl Acad Sci USA* **98**: 8572–8577. doi:10.1073/pnas.141229498
- Shen W, Gorelick RJ, Bambara RA (2011) HIV-1 nucleocapsid protein increases strand transfer recombination by promoting dimeric G-quartet formation. *J Biol Chem* **286**: 29838–29847. doi:10.1074/jbc.M111.262352
- Shrestha P, Shan X, Dhakal S, Tan Z, Mao H (2014) Nascent RNA transcripts facilitate the formation of G-quadruplexes. *Nucleic Acids Res* **42**: 7236–7246. doi:10.1093/nar/gku416
- Sket P, Plavec J (2015) Diversity of DNA and RNA G-quadruplex structures. *Future Science* **2**: 22–36. doi:10.4155/fseh12013.13.28.

- Spindler L, Drevensek-Olenik I, Copic M, Cerar J, Skerjanc J, Mariani P (2004) Dynamic light scattering and  $^{31}\text{P}$  NMR study of the self-assembly of deoxyguanosine 5'-monophosphate: the effect of added salt. *Eur Phys J E, Soft Matter* **13**: 27–33. doi:10.1140/epje/e2004-00037-0
- Sundquist WI, Heaphy S (1993) Evidence for interstrand quadruplex formation in the dimerization of human immunodeficiency virus 1 genomic RNA. *Proc Natl Acad Sci* **90**: 3393–3397. doi:10.1073/pnas.90.8.3393
- Todd AK, Johnston M, Neidle S (2005) Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic Acids Res* **33**: 2901–2907. doi:10.1093/nar/gki553
- Vummi BR, Alzeer J, Luedtke NW (2013) Fluorescent probes for G-quadruplex structures. *Chem Bio Chem* **14**: 540–558. doi:10.1002/cbic.201200612
- Wang Y, Patel DJ (1993) Solution structure of the human telomeric repeat d[AG<sub>3</sub>(T<sub>2</sub>AG<sub>3</sub>)<sub>3</sub>] G-tetraplex. *Structure* **1**: 263–282. doi:10.1016/0969-2126(93)90015-9
- Warner KD, Chen MC, Song W, Strack RL, Thorn A, Jaffrey SR, Ferré-D'Amaré AR (2014) Structural basis for activity of highly efficient RNA mimics of green fluorescent protein. *Nature Struct & Mol Biol* **21**: 658–663. doi:10.1038/nsmb.2865
- Webba da Silva M (2007) NMR methods for studying quadruplex nucleic acids. *Methods* **43**: 264–277. doi:10.1016/j.ymeth.2007.05.007
- Xiong YX, Huang ZS, Tan JH (2015) Targeting G-quadruplex nucleic acids with heterocyclic alkaloids and their derivatives. *Eur J Med Chem* **97**: 538–551. doi:10.1016/j.ejmech.2014.11.021
- Xu S, Li Q, Xiang J, Yang Q, Sun H, Guan A, Wang L (2015) Directly lighting up RNA G-quadruplexes from test tubes to living human cells. *Nucleic Acids Res* **43**: 9575–9586. doi:10.1093/nar/gkv1040
- Xu Y, Kaminaga K, Komiyama M (2008a) Human telomeric RNA in G-quadruplex structure. *Nucleic Acids Symposium Series* **52**: 175–176. doi:10.1093/nass/nrn089
- Xu Y, Kaminaga K, Komiyama M (2008b) G-quadruplex formation by human telomeric repeats-containing RNA in Na<sup>+</sup> solution. *J Am Chem Soc* **130**: 11179–11184. doi:10.1021/ja8031532
- Xu Y, Suzuki Y, Komiyama M (2009) Click chemistry for the identification of G-quadruplex structures: discovery of a DNA-RNA G-quadruplex. *Angewandte Chemie* **48**: 3281–3284. doi:10.1002/anie.200806306
- Xu Y, Ishizuka T, Kimura T, Komiyama M (2010) A U-tetrad stabilizes human telomeric RNA G-quadruplex structure. *J Am Chem Soc* **132**: 7231–7233. doi:10.1021/ja909708a
- Xu Y, Komiyama M (2012) Structure, function and targeting of human telomere RNA. *Methods* **57**: 100–105. doi:10.1016/j.ymeth.2012.02.015
- Zhang N, Phan AT, Patel DJ (2005) (3 + 1) Assembly of three human telomeric repeats into an asymmetric dimeric G-quadruplex. *J Am Chem Soc* **127**: 17277–17285. doi:10.1021/ja0543090
- Zhang S, Wu Y, Zhang W (2014) G-quadruplex structures and their interaction diversity with ligands. *Chem Med Chem* **9**: 899–911. doi:10.1002/cmde.201300566